Designer: Christina Chiddo **Editor:** Lola Olsen

Issue 17
August 2004

CEOS IDN NEWSLETTER

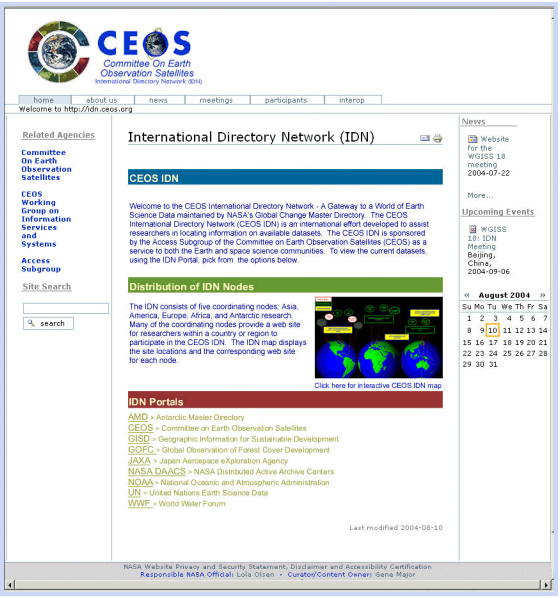
INSIDE THIS ISSUE

Explore the New CEOS IDN Website!

By Rosy Cordova, IDN Software Developer

The CEOS International Directory Network (IDN) web site (http://idn.ceos.org) offers a fresh layout, revised navigation, and the conforming CEOS logo. The web site continues to run on a Zope Application server; however, this server has been enhanced with Plone for improved content management.

Using Plone for the IDN website offers new benefits for individual nodes, such as assuming responsibility for organizing their own sections and adding new content using any web browser running on their desktops. Another benefit of using Plone is that it provides the ability to control the display permissions for web documents. Plone will allow a qualified node representative to set a document's permission to: (1) allow other node participants access to the document, or (2) allow the public or 'anonymous user' unrestricted access. The node representative will only need an account on the website to have access to the maintenance utilities that Plone provides. Documents that can be added to the site include: contact status, upcoming event announcements, news items, presentations for IDN sessions, and metrics.



We are very enthused about the new IDN site, and we hope you will participate. Each node is encouraged to select a representative from their site to participate. Contact Rosy Cordova, from the GCMD staff (cordova@gcmd.nasa.gov), to set up the IDN Node accounts.

For more information on Zope or Plone, check out their websites at:
http://www.zope.org
http://www.plone.org

18th WGISS Plenary and 17th Sub Groups Meeting

By Lola Olsen, NASA, IDN Task Team Lead

Sept. 6, 2004 to Sept. 10, 2004 in Beijing
Deadline for registration was July 15, 2004.

Viktor Pusztai, UNEP/GRID-Budapest will be leading the IDN session in Beijing. All presentations should be coordinated with Viktor for this session: pusztai@mail.kvvm.hu

The IDN session is scheduled for Tuesday afternoon, 7 September 2004 and will be highlighted by the introduction of the all-new CEOS IDN web site (page 1), which now offers the use of Plone. Progress on the new features for MD9, including those related to the spatial query (page 7) and with XPath and SOAP (page 3) will be reported. Also included will be node reports, along with updates on: the use of PostgreSQL for MD9, the modifications for spatial and temporal resolution query, and requirements for model output, the new free-text search using Lucene and the new keyword interfaces (MD9+). Also introduced will be the new mapserver, which emphasizes the increased value of imagery when coupled with vector data. Agenda can be found on page 9.

Viktor Pusztai

Improving Utilization of Earth Observation Satellite Data

By John Faundeen, Working Group on Information Systems and Services (WGISS) Chair, USGS

In response to the Action item related to "Improving Utilization of Earth Observation Satellite Data," the Committee of Earth Observation Satellites (CEOS) members plan to develop information systems with more integrated catalog, search, ordering, and retrieval mechanisms. One of the first steps has been accomplished – that of adding an IDN Portal to the CEOS web site (http://wgiss.ceos.org/), filtered to include data sets from CEOS organizations applicable to atmospheric, oceanography and terrestrial research.

A major effort will be made to identify additional data sets held by CEOS agencies to make the site more complete. The new authoring tool, docBUILDER, will be used to add the new data sets. This tool is available at:
http://gcmd.nasa.gov/User/authoring.html

American Coordinating Node Hosts 3 Summer Interns

By Dennis Petesch, Walter Lowe, and Christina Chiddo

This summer, NASA-sponsored internship programs offered the IDN three amazing college students. These summer internships are designed to offer college students an opportunity to have hands-on experience in a field relating to their major.

Dennis Petesch

Dennis is majoring in Computer Science and Math at Clarke College in Dubuque, Iowa, and is now approaching his senior year. He used the summer to work on the XML Query Language Project. He has researched XPath, a language for addressing parts of an XML document and examined its sufficiency in filtering XML Metadata. Though XML databases were a new concept for him, he proved to gain a strong grasp of the concept.

To finish up his last few weeks of the internship, Dennis plans to research Jaxen, an open source parser that can evaluate XPath expressions among multiple models. In addition, he hopes to learn how to integrate within code to obtain optimal efficiency.

Dennis Petesch was selected through The Visiting Student Enrichment Program (VSEP) from over 260 applicants for an 8-week internship position at GSFC. VSEP offers students opportunities to work with professionals at a world-class facility while experiencing work through a project primarily focused on computer science or the application of computers to solve problems in other sciences.

Walter Lowe

Walter, a Computer Science major at the University of Maryland spent his summer internship with the Global Change Master Directory project.

"My work over the past few weeks has been to implement a new remote interface for the GCMD. To do this, I have been using Simple Object Access Protocol (SOAP) and Web Services Description Language (WSDL) technologies. It is a tool used to send messages and commands back and forth between computers over a network. SOAP is useful because it runs over the Internet and communicates using the universal language XML.

In order to inform computers on how to communicate with the GCMD, I have been working with Web Services Description Language (WSDL). This tool is used to specify the types of commands that the GCMD will allow and how to issue these commands. WSDL is especially useful because it allows users to create client programs using the programming language of their choice.

The first two weeks of my internship were spent learning about various web technologies. During this time, I learned about the various tools that I would be using to complete my task. Afterwards, I began implementation of the SOAP server system. As I continue on the project, I hope to complete the implementation of the server. Additionally, I will draft documentation which will allow users to easily take advantage of the new SOAP server."

Eleanor Branch

Eleanor is from Stanford University and is majoring in Earth Systems/Energy. Approaching her senior year, she dedicated her first few weeks of the internship to review and update all preexisting energy related data set and service descriptions. She created 20 new DIFs and SERFs on renewable energy topics such as Ocean Tidal & Wave Energy, Offshore Wind Power Resources, U.S. Nuclear Waste Storage, and Lightning Phenomena.

Throughout her last weeks, Eleanor helped with a new design for a keyword search on the GCMD and created two new Learning Center pages for Energy Resources, which will be available from the GCMD Learning Center at a later date.

Walter Lowe and **Eleanor Branch** were selected by the GSFC/Howard University Fellowship in Atmospheric Science (GoHFAS) partnership for their summer internship. GoHFAS's goal is to increase the number of students from traditionally underrepresented groups that pursue terminal degrees in atmospheric science and related fields. Pursuant to these goals, their objectives are to expand participation by Howard faculty and students in NASA Atmospheric Science related research and to provide the Howard University Program in Atmospheric Science (HUPAS) with a pool of applicants with sufficient undergraduate exposure in the discipline.

Usability Testing of the GCMD

By Monica Holland, IDN, Ocean Sciences Coordinator

Inspired by the feedback reported by the IDN's Canadian Centre for Remote Sensing (CCRS) usability study, the American Coordinating Node recently conducted usability testing to gain valuable insights from a user's perspective. Usability is defined as the effectiveness, efficiency, and satisfaction with which users can achieve tasks in a particular environment of a product.

According to several studies, "Usability, and the user experience, should be considered at every step during a system's life cycle. From conceptualization and requirements definition, through prototyping, development, product release, and even at post-deployment, focusing on a product's usability will result in lower lifetime development costs and increased user productivity and satisfaction."

Usage peaked after the release of the new web design in January 2004 and has remained steady. In an effort to consistently offer accessible content and a functional site design, the American Coordinating node was evaluated by six reviewers (including a college science professor, a research professional, two science students and two computer science students).

The comments received from the usability testing have proved useful. A few interesting comments that emerged from the usability testing include:

A) "Why is the 'Update Record' button red? It grabs attention away from other buttons." (DIF record)."

B) "Cluttered/Intimidating, confused by the small words" used as anchor tags on the Keyword search pages.
C) Liked the idea that the left hand navigation was the same as the topics from the home page (navigation).
D) "Well organized, pictures help to explain topics" (layout of site)
E) "Straight forward heading/title" (design of title image).

Plans to update the existing web pages to assure ease of use and overall functionality include:

(1) Re-design Keyword page display; remove "Go to" anchor links at top of page.
(2) Update Navigation "tabs"; summarize information available under each tab with alternate text tags and modify nomenclature of tabs.
(3) Enhance the Data Services page by including images.
(4) Enhance other features on web site (e.g., make FAQ pages easier to locate).
(5) Enhance the layout of the DIF record.

The Antarctic Data Center of Argentina (CDA)

By Celia E. Izquierdo, CDA, Manager
Translated by Cheryl Solomon, USGS/BRD Biological and Ecosystems Coordinator

El Centro de Datos Antárticos de Argentina (CDA) es un usuario continuo de las herramientas y nuevas utilidades que brinda el AMD y GCMD. Actualmente y gracias a la colaboración de ambos en su diseño e implementación, el CDA cuenta con un Portal permite mostrar su metadata. A través de este, se realiza la incorporación de nuevos metadatos, sus actualizaciones y todo tipo de consultas.

Referente a la herramienta en uso, esta nueva versión cuenta con facilidades de acortan notablemente el tiempo necesario para la redacción de DIF, siendo de suma utilidad los archivos de datos de investigadores, personal de los Centros, Nombres de Proyectos, etc. que dinamizan fuertemente la tarea facilitando la inclusión de información sin necesidad de reescribirla.

De igual manera el CDA siempre ha recibido una inmediata respuesta el personal ad hoc del AMD - GCMD, ante cual duda o necesidad surgida en su tarea.

The Antarctic Data Center of Argentina (CDA) is a constant user of the tools and new services which the Antarctic Master Directory and GCMD offer. At the present time, and thanks to the collaboration of both AMD and GCMD in its design and implementation, the CDA relies on a portal which allows it to display its metadata. Through this portal, the incorporation of new metadata, its updates, and all types of searches are possible.

Referring to the latest tool, this new version includes services which notably reduce the time necessary to write a DIF, being of utmost utility the data archives of investigators, Centers' personnel, names of projects, etc., which make the task highly dynamic, facilitating the inclusion of information without having to rewrite it.

Similarly, CDA has always received an immediate response from AMD-GCMD ad hoc personnel, whenever any questions or necessities arise during our work.

Science User Working Group Meeting Held on May 25th and 26th, 2004

By Martin Ruzek, Acting Chairman Science Users' Working Group,
Universities Space Research Association

The Science Users' Working Group met on May 25th and 26th, 2004 — just two weeks after the CEOS meeting in Tromso, Norway — to provide guidance for the future of the American Coordinating Node of the CEOS IDN (the GCMD).

The GCMD reported on its progress during the year, and the UWG members were all impressed with the level of activity and continuing community benefit of GCMD. The GCMD continues to strive for efficiency and usability, as evidenced by the evolution of the MD software, development of the docBUILDER DIF authoring tool, redesign of the web front end to GCMD, growth and maintenance of the DIF and SERF collection, keyword vocabulary updates, etc.

The UWG made several broad recommendations which will be further refined in its report to GCMD:

- Develop a strategic plan built around NASA's guiding Earth science and application themes.
- In light of potential (now implemented) NASA organizational changes, consider including space science data and user services.

(continued on next page)

(continued from page 5)

- Continue to explore ways to better understand who the users of the IDN are, and direct efforts to meet defined user needs.
- Pursue and advance cataloging and metadata collaborations with NOAA and NSF Cyberinfrastructure programs such as NMMR, THREDDS, NSDL, and DLESE.
- Encourage new science programs to agree to submit metadata to IDN as part of their program plans.
- Establish the GCMD's role as keyword-controlled vocabulary broker.
- Allow search and indexing robots access to the IDN collection in a way that does not compromise system performance.

The UWG looks forward to working with the IDN to further refine these recommendations.

Note:

The GCMD team has already responded to several recommendations by creating a separate cache of data set descriptions — specifically for search engine robots to index (a recommendation also heard at the IDN session in early May) and by conducting external web site usability testing.

GO-ESSP Conference (formally ESP) and Progress

By Jason Divock, IDN, Software Developer

Jason Divock and Lola Olsen attended the third annual Global Organization for Earth System Science Portal (GO-ESSP) meeting in Princeton, New Jersey on June 8th and 9th. One of GO-ESSP's main goals is to develop a new generation of software infrastructure to provide distributed access to observed and simulated data from the climate and weather communities. To attain this goal, the group plans to integrate individual software components - building a federation of frameworks that can work together using consensus standards.

The conference offered a wide variety of topics that included both groups assembling and coordinating data and others who were pioneering technologies to make these tasks easier and more powerful.

Those participants who presented demos of their progress were offered constructive feedback and possible new directions for their work. At the end of the conference, a number of 'main topics' were selected and a point of contact for each was established. Two of the main topics identified will be represented by a model metadata group and a middleware group. This was done to help insure that interest in the topics did not fade after the conference but were kept on the forefront and in discussion.

The progress of the IDN was presented, with new partners identified for collaborations. The conference was a terrific way for the GCMD to stay current with the progress of others in the community and look for inspiration for development ideas.

New Search Capabilities for IDN Data

By Mark Harwood, InPerspective Technology

An interest in the improvement of the spatial search for Earth science data prompted IDN representatives to contact InPerspective, our new company focused on these perceived needs of the community. Mutual interests have resulted in a collaboration to improve the search for the IDN's spatial data and include:

- More relevant search results
- Better support for defining locations of interest
- New advanced text search options
- Improved search results display showing relevant extracts from matching documents
- Automated suggestions of related words that can be used to refine queries.

The technology which offers these features is based on the Apache foundation's "Lucene" software, which is a powerful search library suitable for both novice and sophisticated searchers. For the average user who performs basic queries by typing in a number of keywords, the search engine is capable of returning the most relevant results by using statistical natural language processing techniques to identify the most significant terms from the query and underlying documents. For power-users who wish to have more control over queries, the query syntax supports powerful search features such as Boolean, phrase, fuzzy, wildcard and fielded queries as well as term and document boosting.

The Lucene search software has been extended to add support for spatial queries using the OpenGIS standards defined for spatial databases. Documents held in the search index can define locations using a set of coordinates that express coverage as a shape e.g. a polygon, point, line, multi-point, or multi-polygon. Queries also use such shapes to define the user's area of interest, and these subsequently select matching documents using spatial comparison functions such as "intersects", "contains" or "crosses". This more sophisticated search capability creates the interesting possibility of allowing DIF files to define more precisely the spatial coverage they represent; for example, by using one or more polygons or a set of discrete points.

Of course users rarely want to type coordinates in to perform a search; therefore, work remains in putting a new mapping system in place to let users select their area of interest from a map.

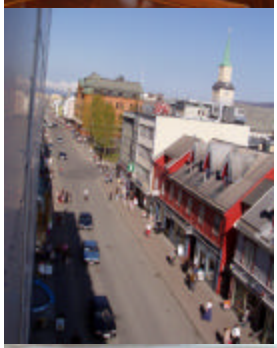
A prototype demonstration of the new search technology is available at www.inperspective.com.

INPERSPECTIVE

IDN Task Team Notes

By Rosy Cordova, IDN Software Developer, & Lola Olsen, NASA, GCMD Project Manager

May 2004, Tromso, Norway



The IDN task team meeting was coordinated by Lola Olsen, the task team lead, beginning with a review of the minutes from the last task team meeting at Chiang Mai, Thailand. Lola reviewed the IDN reports and the work presented by the American Coordinating node at that time. (http://idn.ceos.org/IDN/Meetings/2003_09_Chiang_Mai/minutes.html)

The review of the metrics of content in the IDN was given from January 1999 to April 2004. Data set descriptions, DIFs, were analyzed by Earth Science Topic (Atmosphere and Biosphere the highest at 16% each) and IDN Node (NASA highest at 44%). Then the website usage metrics were presented. The highest number of users connecting to the site have numeric (25%) and foreign IPs (23%).

The progress of the MD9 package was then evaluated. The MD9 setup was presented. It included the data flow model, the client/server setup, the query language for an open search interface, and the LDA for distributed input from IDN nodes. Other upgrades that were reviewed were the new keyword interface, the subscription service, the auto notification service, the geospatial search, and docBUILDER.

The IDN nodes present then gave their status reports. The nodes were: CGDI, CNES, DLR, ESA, JAXA, JCADM, and UNEP/GRID. Other IDN representatives led discussions on agenda items. For example, Lorant Czaran (UNEP/GRID) reviewed the status of the ISO modifications and what the IDN has done to comply with ISO.

The future plans for the IDN associated with the timelines for future MD releases were also reviewed. These included evaluations/requirements/testing, such as: (1) evaluation of spatial and temporal resolution proposal by Ivan Petiteville (ESA); (2) requirements for model output by Glenn Rutledge (NOAA/NCDC); (3) testing of PostgreSQL by UNEP/GRID-Budapest (Victor Pusztai); (4) docBUILDER review by John Faundeen (USGS); (5) evaluation of the Joint Committee on Antarctic Data Management's (JCADM) portal by Stein Tronstad of the Norwegian Polar Institute; and (6) evaluation of the keywords for use by CIP (Jolyon Martin of ESA).

To view all the presentations given at Tromso, please go to the IDN website (http://idn.ceos.org/IDN/Meetings/2004_05_Tromso/). Many thanks to the organizers of the Tromso meeting.

The IDN Agenda for Beijing Meeting

Tuesday, September 7, 2004

2:00 Introduction by Viktor Pusztai
Review minutes from Tromso

2:05 New CEOS IDN Web Site
Demonstrate use of site and receive comments.

2:20 Node Reports

2:50 New Interfaces
Progress with SOAP
Progress with XPATH

3:10 XSLT FGDC-to-DIF translator

3:20 Progress on the use of PostgreSQL for MD9

3:30 Status of modification for spatial and temporal resolution query

3:40 New "input" for model "output"

3:45 New keywords since last CEOS meeting

3:50 New location, instrument, and platform hierarchies for improved search

4:00 Update on Lucene
- In combination with keyword search
- New spatial query

4:20 New Mapserver

4:30 Services will be presented in Services Task Team

CEOS
Committee On Earth Observation Satellites
International Directory Network (IDN)